Identification of Disease Genes

This exercise focuses on the identification of a disease gene using NCBI's human genome assembly. The reference human genome assembly along with integrated maps, literature, and expression information comprises a powerful discovery system for exploring candidate human disease genes.

Problem: A laboratory has generated an EST library from a hemochromatosis patient and wants to identify the gene(s) causing the phenotype.

We will follow these steps to solve the problem:

1. Compare an EST to the human genome (using BLAST).
2. Identify the gene(s) aligning with the EST and download their sequences (using MapViewer).
3. Identify whether the EST contains any known SNPs (using dbSNP).
4. Determine whether a mutant form of the gene causes a phenotype (using OMIM).

Problem 1:

A laboratory has generated an EST library from a hemochromatosis patient and wants to identify the gene(s) causing the phenotype.

Outline:

We will follow these steps to solve the problem:
1. Compare the EST from a hemochromatosis patient to the human genome (using BLAST).
2. Identify the gene(s) aligning with the ESTs and download their sequences (using Map Viewer).
3. Identify whether the EST contains any known nucleotide variations (single nucleotide polymorphisms) (using dbSNP).
4. Determine whether a mutant form of the gene is known to cause a phenotype (using OMIM).

Step 1. Compare ESTs to the human genome (using BLAST):

One way to identify the genes expressing the EST is to compare the EST sequence with the human genome assembly and the genes annotated on it. To

access the specialized BLAST page for searching against the human genome assembly, click on

BLAST (human genome)

Copy the EST sequence provided below. Paste it in the query box of the BLAST page and select the “genome (reference only)” database. Start the search by clicking on the “Begin Search” button.

Query EST Sequence:

TGCCTCCTTTGGTGAAGGTGACACATCATGTGACCTCTTCAGTGACCACTCT
ACGGTGTCGGGCCTTGAACTACTACCCCCAGAACATCACCATGAAGTGGCT
GAAGGATAAGCAGCCAATGGATGCCAAGGAGTTCGAACCTAAAGACGTATT
GCCCAATGGGGATGGGACCTACCAGGGCTGGATAACCTTGGCTGTACCCC
CTGGGGAAGAGCAGAGATATACGTACCAGGTGGAGCACCCAGGCCTGGAT
CAGCCCCTCATTGTGATCTGGG

Name the chromosome and the contig that we get as a BLAST hit. Is the EST sequence 100% identical to the genomic sequence? Note the nucleotide difference and position between the two sequences.

Step 2. Identify the gene(s) expressing the ESTs and download their sequences:

To visualize the BLAST hit on the genome using Map Viewer, click on the "Genome View" button at the top of the results page, then on the Map element "NT_007592". Currently, 4 maps should be displayed (Contig, Model, RNA and Gene_seq). Zoom out 2 or 4 times by clicking on right most contig map and selecting the appropriate option.
The BLAST hit, indicated by the red bar, is in the region of one of the exons of the HFE gene annotated on the human genome. Make the Gene_seq map a master map by clicking on the arrow at the top of the map. Display the entire HFE gene sequence by clicking on the "dl" link and then on "Display". Copy the sequence and paste it in the area provided below. We will use it later to obtain the exon-intron structure. You can adjust the nucleotide locations to download the upstream or downstream sequence by using the "adjust by" and "Change Region/Strand" option.

Step 3. Determine whether the ESTs contain known SNPs:

Go back to the Map Viewer report. Click on the Maps and Options link. Remove all the maps except the Gene_seq map by selecting the map under the Maps Displayed menu and clicking on Remove. Now add the variation map from the Available maps menu (by selecting the map and clicking on Add). Make the Variation map as the master map by selecting it and clicking the Make

Master/Move to Bottom option. Then click on Apply. Now two maps are displayed, Variation (it's the rightmost and master map) and Gene_seq. The master map provides detailed information for the map features, in this case SNPs. ". (The Mini-Course Map Viewer Quick Start describes the usage of the Map Viewer in detail.) Zoom in on the blast hit area (red bar). There are two SNPs in the area, one of them is rs1800562. Click on the link for the SNP. There is an A/G SNP is at the nucleotide position 26033141 on the contig NT_007592 as mentioned under Fasta sequence and Integrated maps. Is this the same nucleotide variation found in the BLAST result in Step 1? Please note that the SNP results in the Cysteine 282 Tyrosine mutation for the longest protein (expressed by the mRNA NM_000410) as reported under GeneView.

Step 4. Determine whether the mutant HFE gene causes a phenotype:

Go back to the Map Viewer report and add the Phenotype map as the master map using the Maps and Options menu. Click on the HFE link that leads to the hemochromatosis record (235200) in the OMIM database. The record details how mutations in the HFE gene are associated with a phenotype, hemochromatosis. Click on the Allelic Variant "View list" to get information about mutant proteins from patients. Is Cys282Tyr variant mentioned in the list? Which phenotype does it cause?

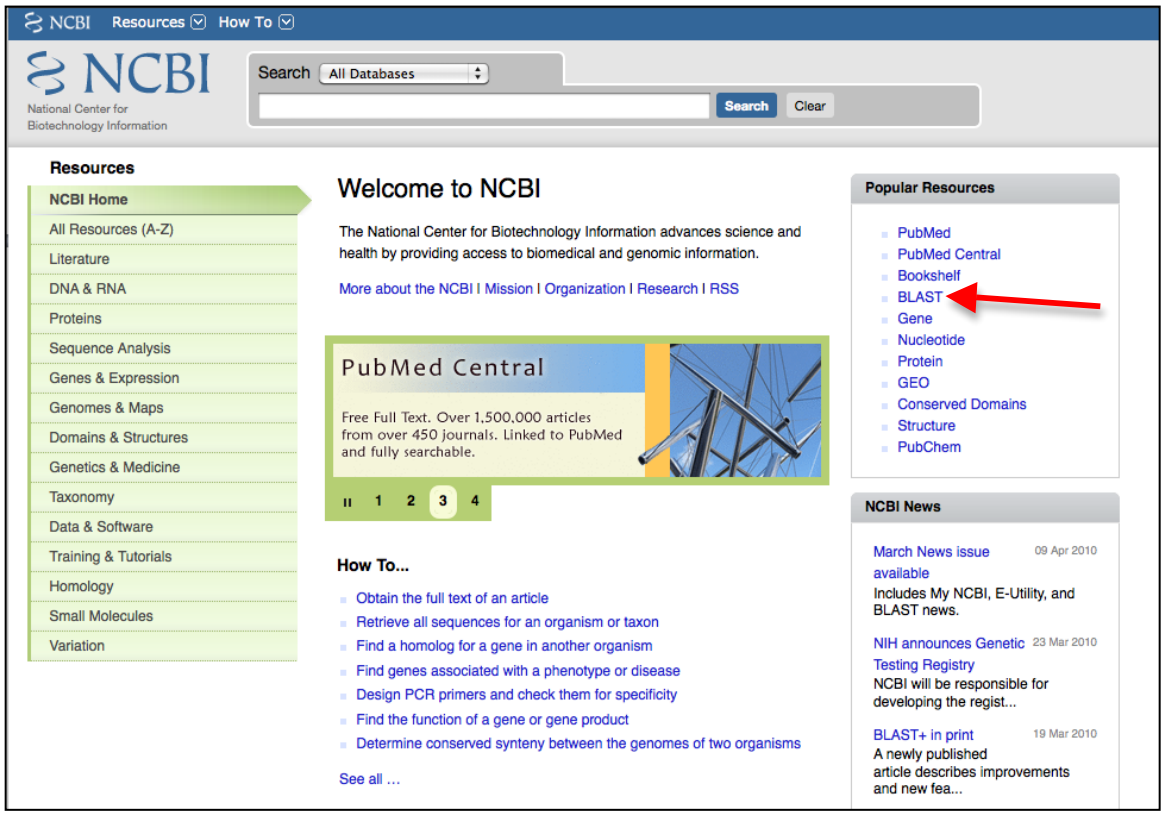
Summary:

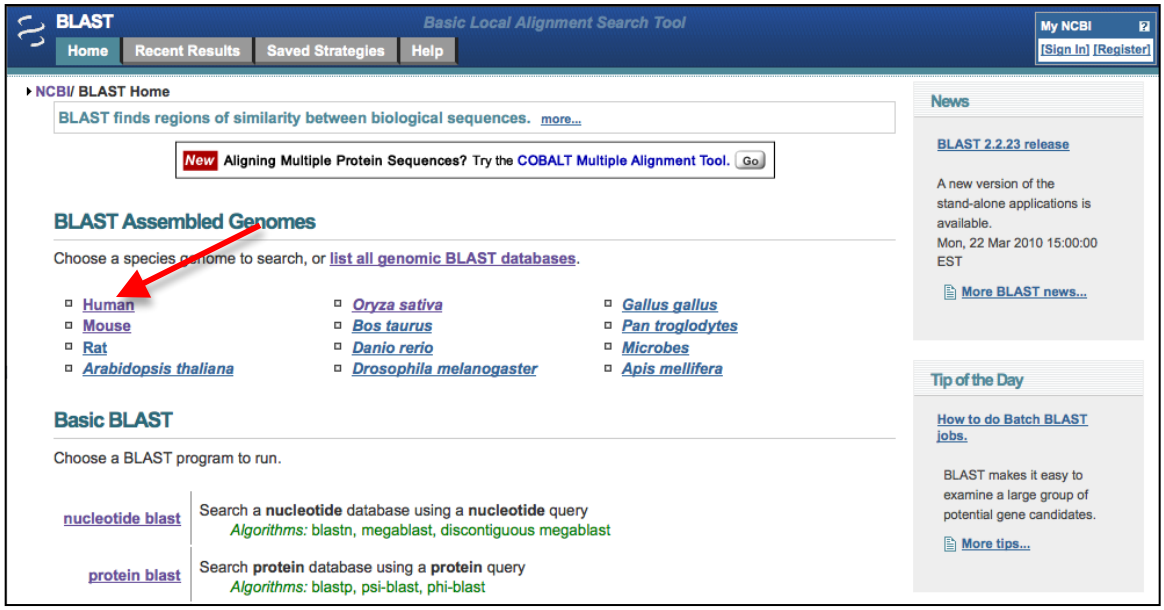
This exercise describes steps to identify the gene expressing an EST obtained from a hemochromatosis patient, download the gene sequence, identify known SNPs in the gene and find SNP-associated phenotypes.

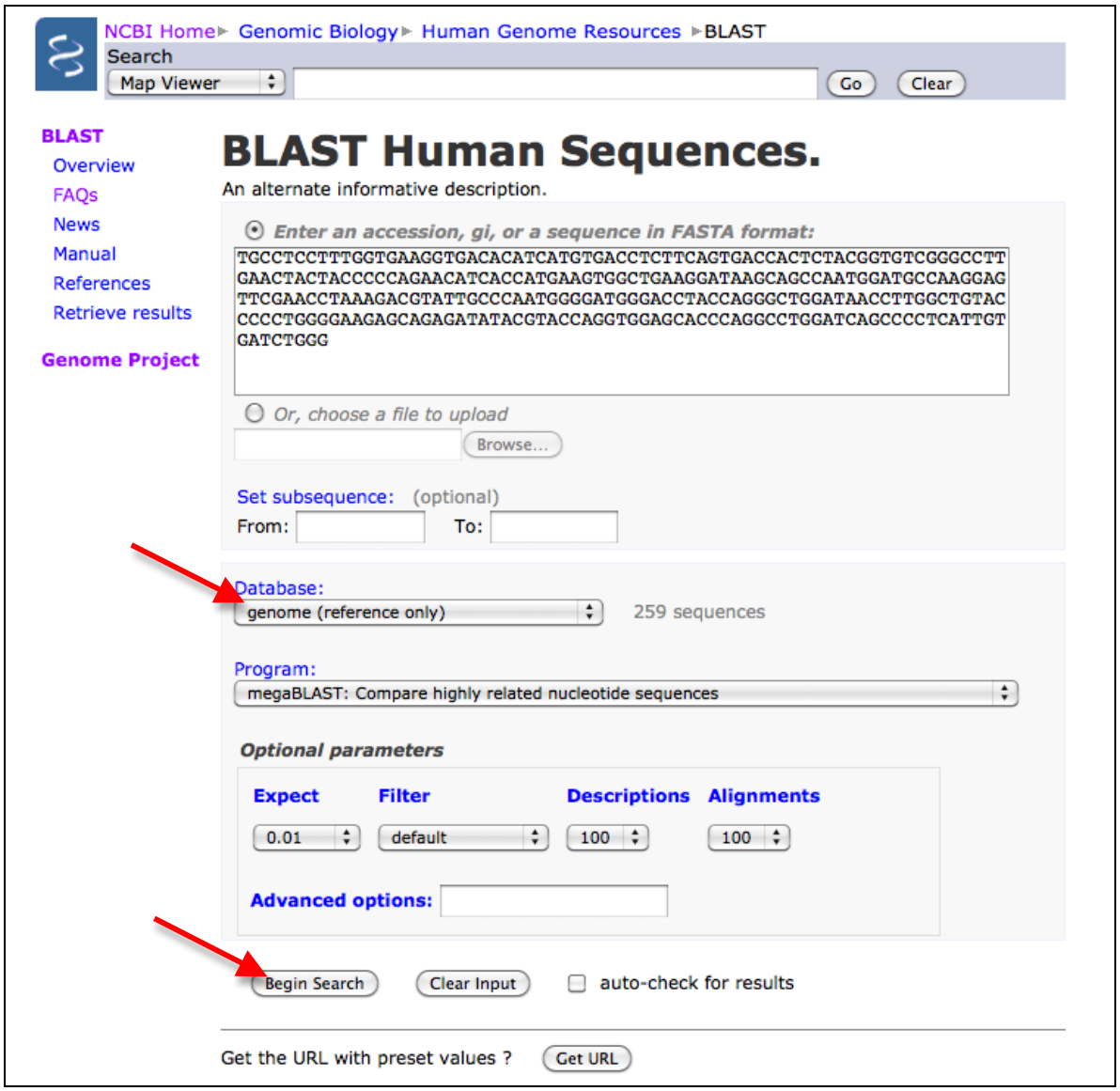
Step 1: The query EST sequence was found to align contig NT_007592.15 on chromosome 6 with one nucleotide difference (G to A with respect to the nucleotide 26033141 on the contig).
Step 2: The query EST was found to be expressed by the HFE gene.
Step 3: The query EST sequence contains a known SNP (G/A with respect to nucleotide 16951392 on contig NT_007592.15).
Step 4: Mutations in the HFE gene are associated with hemochromatosis.

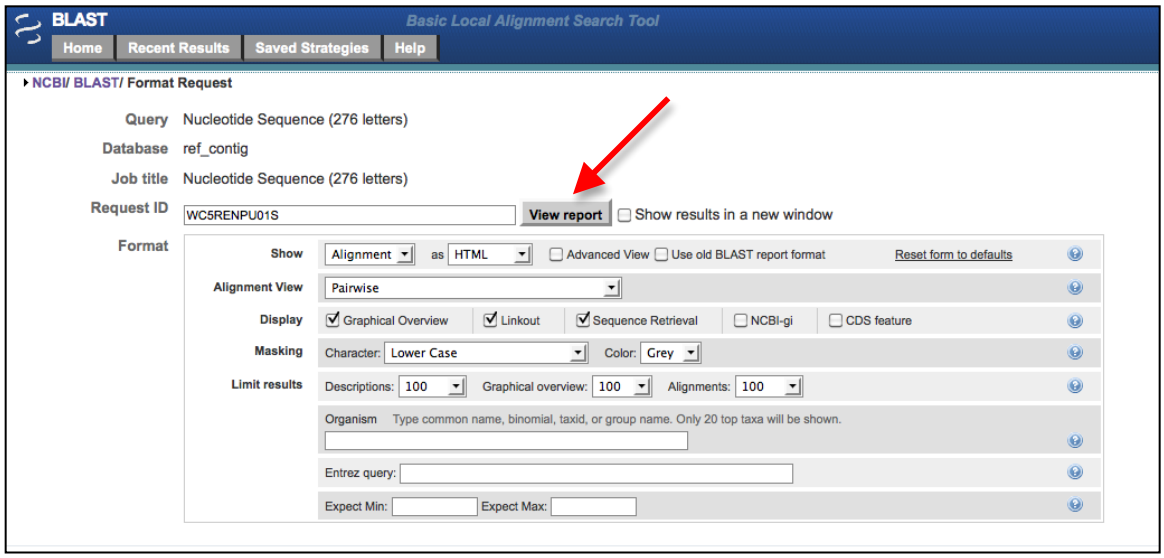
Screen images

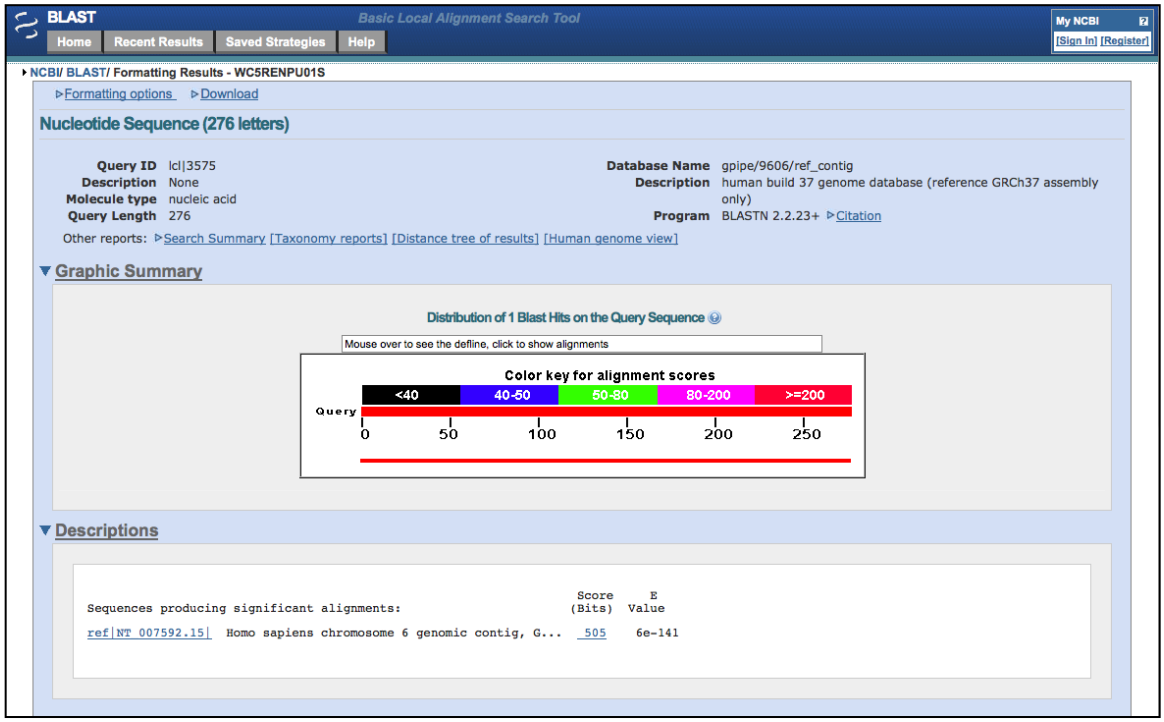
Step 1: Compare EST against the human genome.

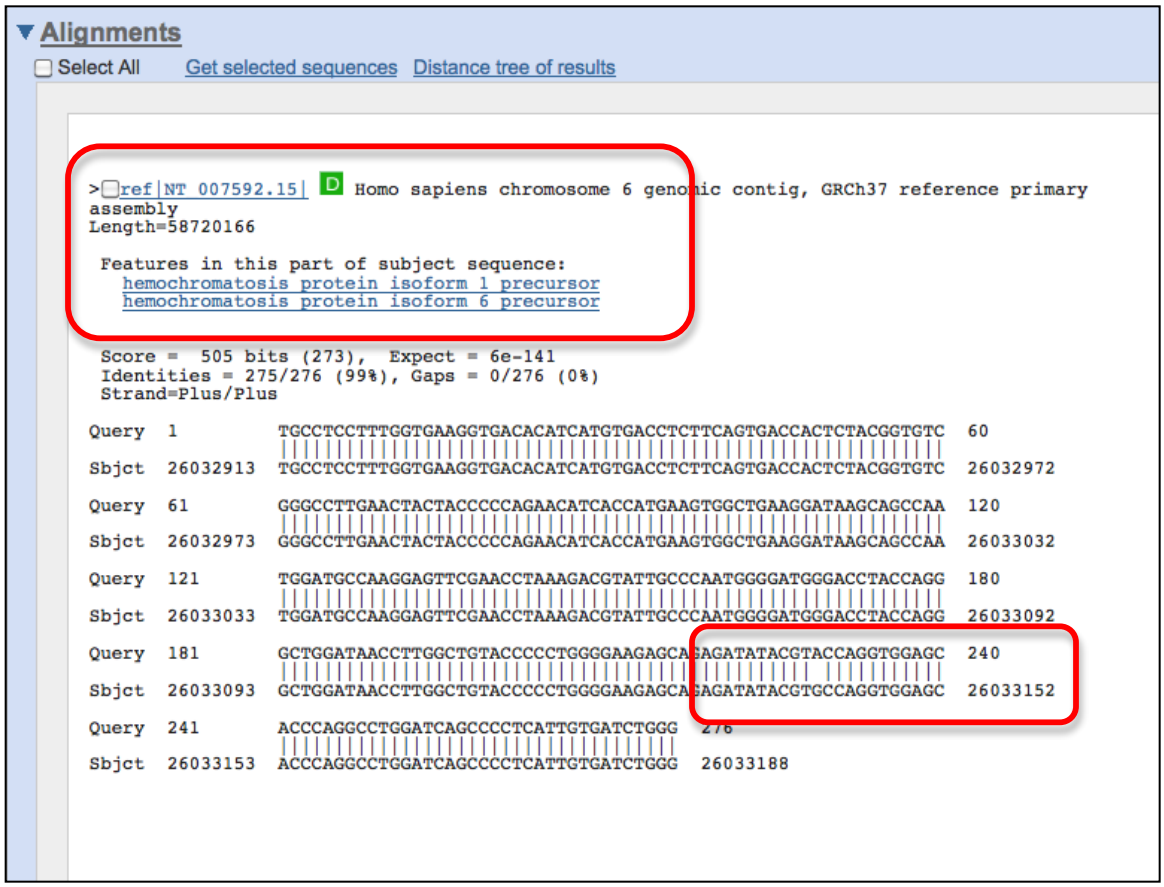





NCBI Home
Genomic Biology
Human Genome Resources
BLAST
Search
Map Viewer
Go
Clear
BLAST
Overview
FAQs
News
Manual
References
Retrieve results
Genome Project
BLAST Human Sequences.
An alternate informative description.
Enter an accession, gi, or a sequence in FASTA format:
TGCCTCCTTTGGTGAAGGTGACACATCATGTGACCTCTTCAGTGACCACTCTACGGTGTCGGGCCTT
GAACTACTACCCCCAGAACATCACCATGAAGTGGCTGAAGGATAAGCAGCCAATGGATGCCAAGGAG
TTCGAACCTAAAGACGTATTGCCCAATGGGGATGGGACCTACCAGGGCTGGATAACCTTGGCTGTAC
CCCCTGGGGAAGAGCAGAGATATACGTACCAGGTGGAGCACCCAGGCCTGGATCAGCCCCTCATTGT
GATCTGGG
Or, choose a file to upload
Browse...
Set subsequence: (optional)
From:
To:
Database:
genome (reference only)
259 sequences
Program:
megaBLAST: Compare highly related nucleotide sequences
Optional parameters
Expect
Filter
Descriptions
Alignments
0.01
default
100
100
Advanced options:
Begin Search
Clear Input
auto-check for results
Get the URL with preset values ?
Get URL

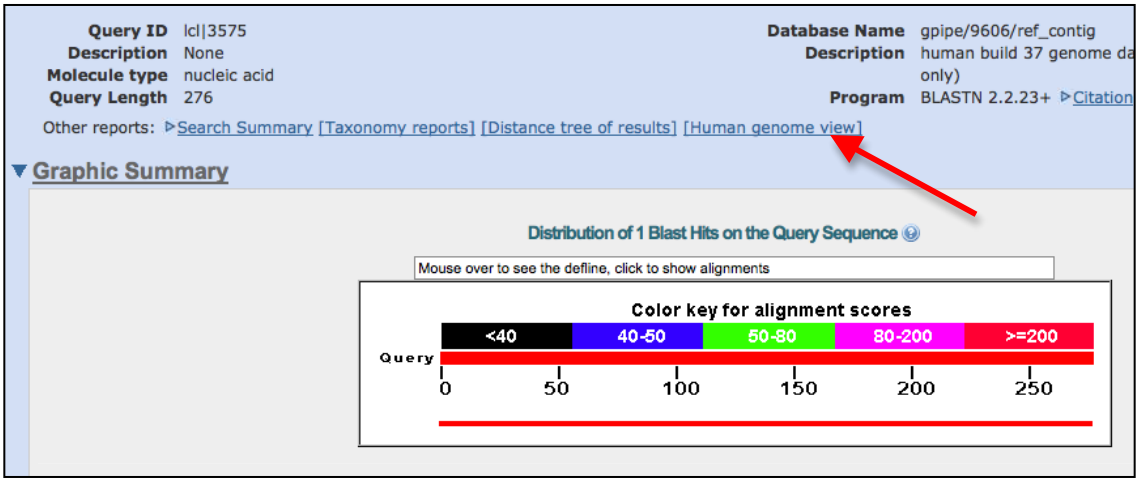

BLAST
Basic Local Alignment Search Tool
Home
Recent Results
Saved Strategies
Help
NCBI/ BLAST/ Format Request
Query Nucleotide Sequence (276 letters)
Database ref_contig
Job title Nucleotide Sequence (276 letters)
Request ID WC5RENPU01S
View report
Show results in a new window
Format
Show Alignment as HTML Advanced View Use old BLAST report format Reset form to defaults
Alignment View Pairwise
Display Graphical Overview Linkout Sequence Retrieval NCBI-gi CDS feature
Masking Character: Lower Case Color: Grey
Limit results Descriptions: 100 Graphical overview: 100 Alignments: 100
Organism Type common name, binomial, taxid, or group name. Only 20 top taxa will be shown.
Entrez query:
Expect Min: Expect Max:

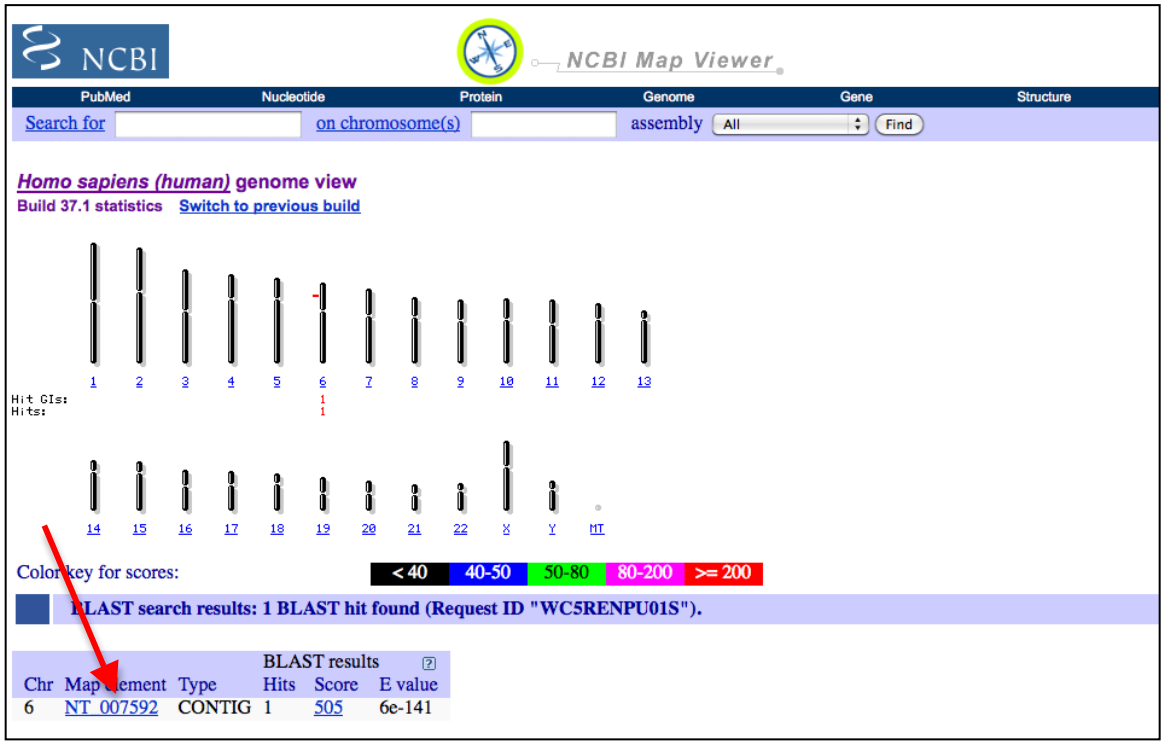

BLAST
Basic Local Alignment Search Tool
Home
Recent Results
Saved Strategies
Help
My NCBI
[Sign In] [Register]
NCBI/ BLAST/ Formatting Results - WC5RENPU01S
Formatting options
Download
Nucleotide Sequence (276 letters)
Query ID lcl|3575
Description None
Molecule type nucleic acid
Query Length 276
Database Name gpipe/9606/ref_contig
Description human build 37 genome database (reference GRCh37 assembly only)
Program BLASTN 2.2.23+ Citation
Other reports: Search Summary [Taxonomy reports] [Distance tree of results] [Human genome view]
Graphic Summary
Distribution of 1 Blast Hits on the Query Sequence
Mouse over to see the defline, click to show alignments
Color key for alignment scores
<40
40-50
50-80
80-200
>=200
Query
0
50
100
150
200
250
Descriptions
Sequences producing significant alignments: Score (Bits) E Value
ref|NT_007592.15| Homo sapiens chromosome 6 genomic contig, G... 505 6e-141

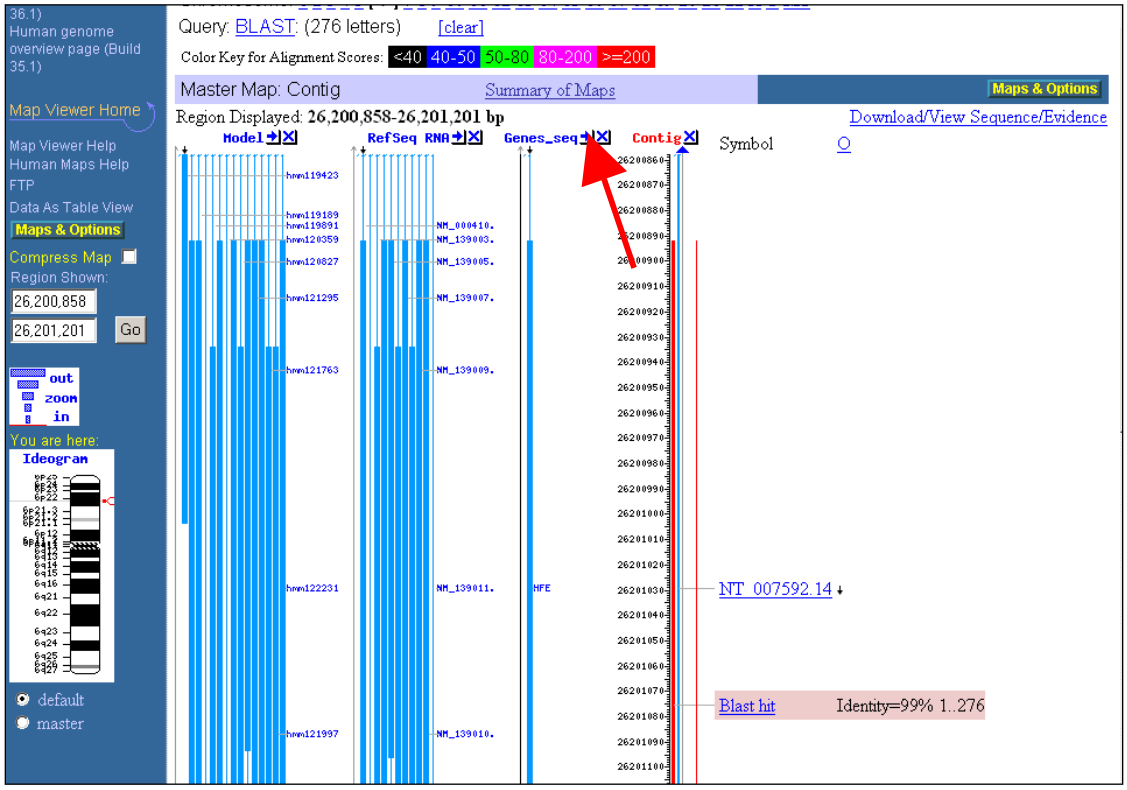


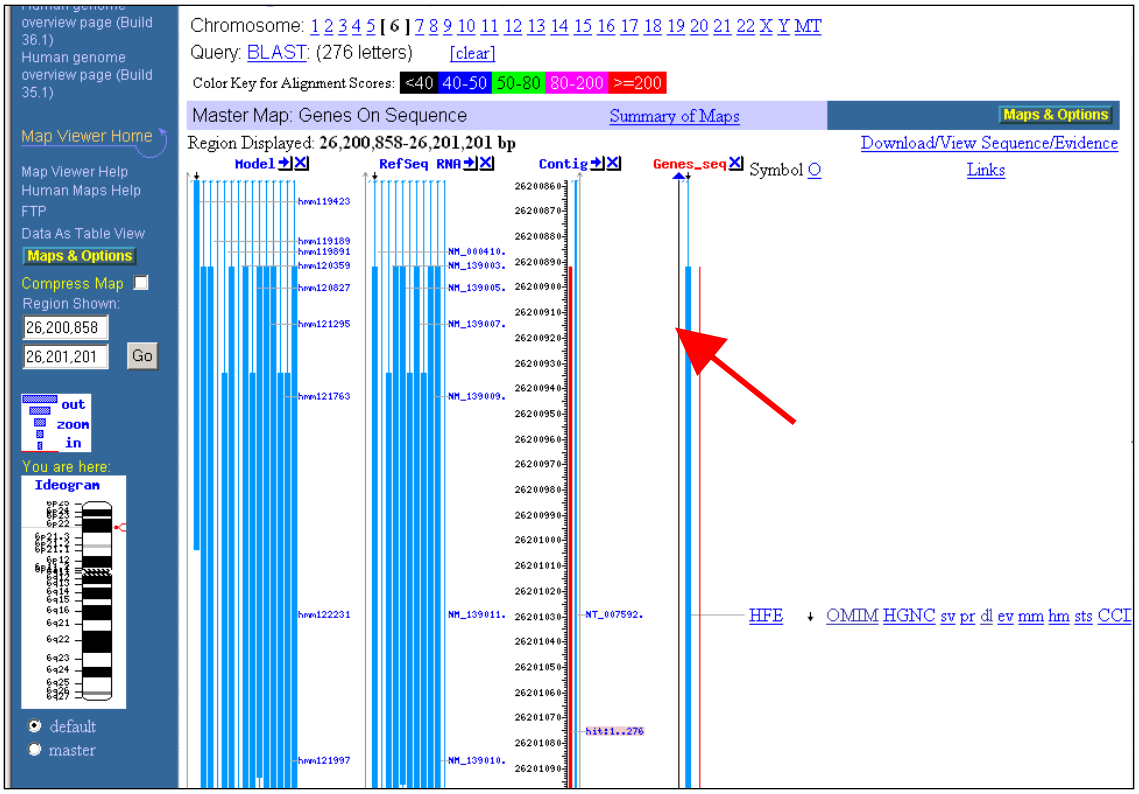
Result: The EST sequence is aligned to the contig NT_007592.15 on chromosome 6 with one nucleotide difference (G to A with respect to the nucleotide 26033141 on the contig).

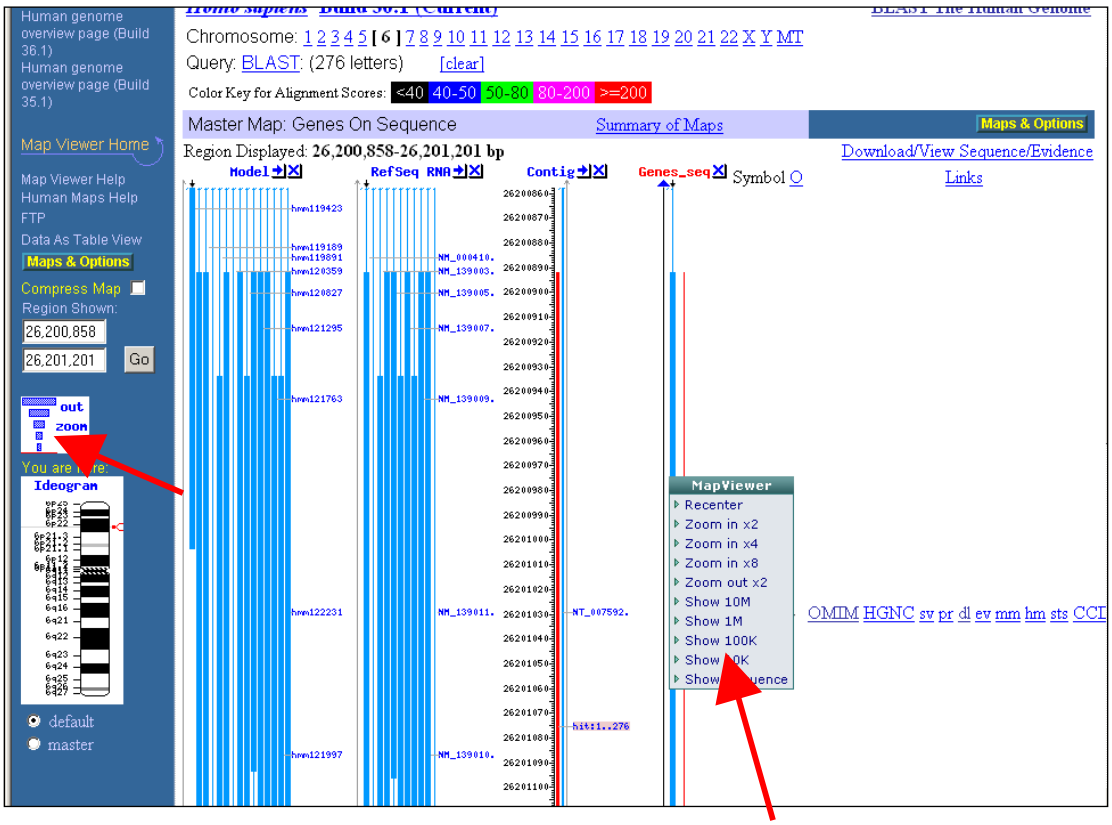
Step 2: Identify the gene(s) expressing the EST and download their sequences

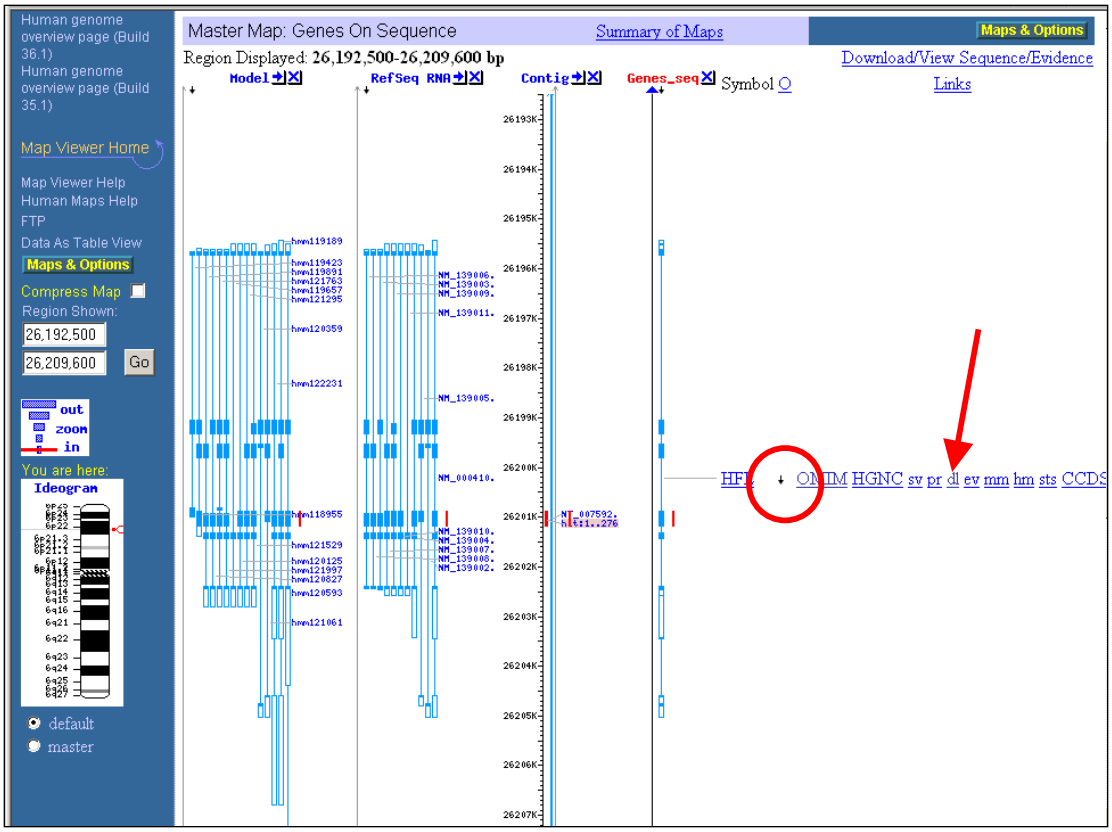


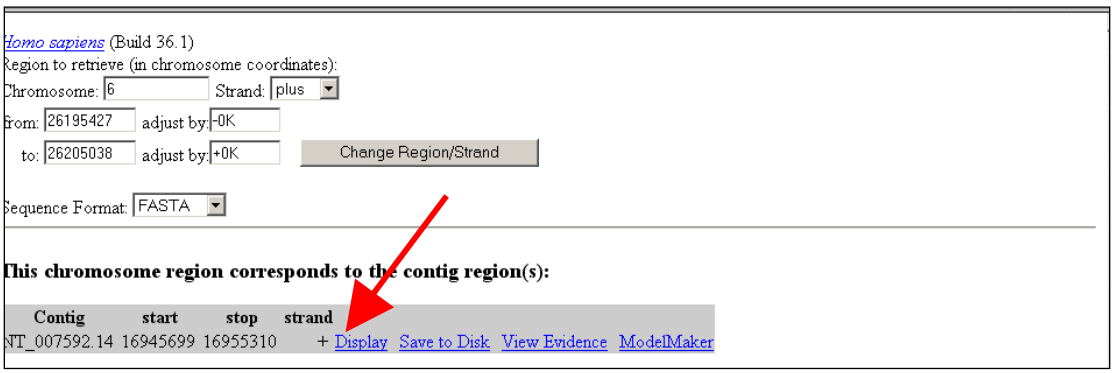

NCBI
NCBI Map Viewer
PubMed
Nucleotide
Protein
Genome
Gene
Structure
Search for
on chromosome(s)
assembly
All
Find
Homo sapiens (human) genome view
Build 37.1 statistics
Switch to previous build
1 2 3 4 5 6 7 8 9 10 11 12 13
Hit GIs:
Hits:
1
1
14 15 16 17 18 19 20 21 22 X Y MT
Color key for scores:
< 40
40-50
50-80
80-200
>= 200
BLAST search results: 1 BLAST hit found (Request ID "WC5RENPU01S").
BLAST results
Chr
Map Element
Type
Hits
Score
E value
6
NT_007592
CONTIG
1
505
6e-141

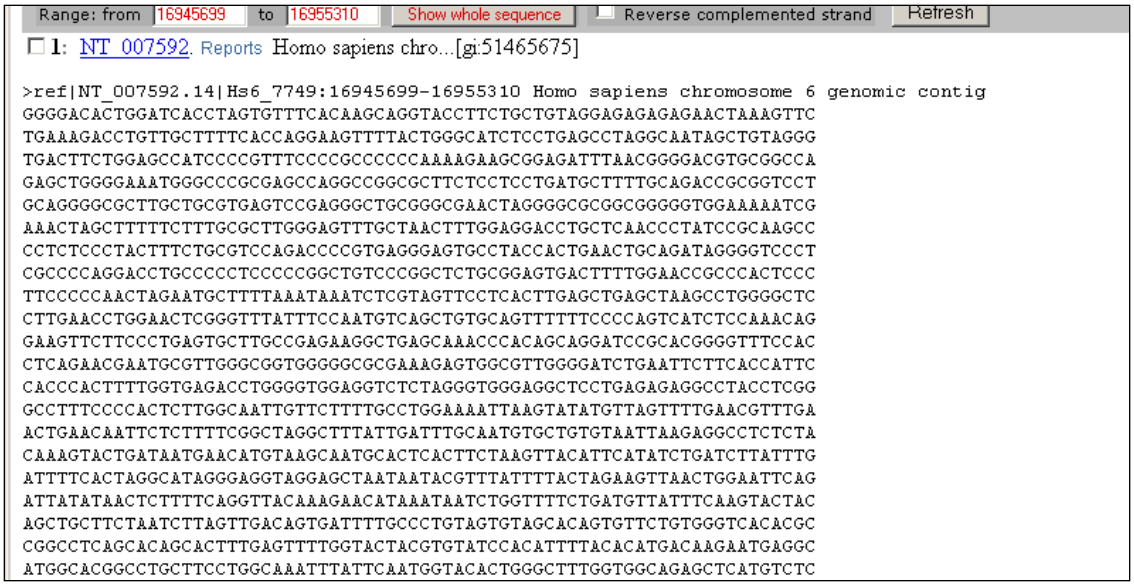

36.1)
Human genome overview page (Build 35.1)
Map Viewer Home
Map Viewer Help
Human Maps Help
FTP
Data As Table View
Maps & Options
Compress Map
Region Shown:
26,200,858
26,201,201
Go
out
zoom
in
You are here:
Ideogram
default
master
Query: BLAST: (276 letters)
[clear]
Color Key for Alignment Scores:
<40
40-50
50-80
80-200
>=200
Master Map: Contig
Summary of Maps
Maps & Options
Region Displayed: 26,200,858-26,201,201 bp
Download/View Sequence/Evidence
Model
RefSeq RNA
Genes_seq
Contig
Symbol
hmm119423
hmm119189
hmm119891
hmm120359
hmm120827
hmm121295
hmm121763
hmm122231
hmm121997
NM_000410.
NM_139003.
NM_139005.
NM_139007.
NM_139009.
NM_139011.
NM_139010.
HFE
NT_007592.14
Blast hit
Identity=99% 1..276


Human genome overview page (Build 36.1)
Human genome overview page (Build 35.1)
Map Viewer Home
Map Viewer Help
Human Maps Help
FTP
Data As Table View
Maps & Options
Compress Map
Region Shown:
26,200,858
26,201,201
Go
out
zoom in
You are here:
Ideogram
default
master
Chromosome: 1 2 3 4 5 [6] 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X Y MT
Query: BLAST: (276 letters) [clear]
Color Key for Alignment Scores: <40 40-50 50-80 80-200 >=200
Master Map: Genes On Sequence
Summary of Maps
Maps & Options
Region Displayed: 26,200,858-26,201,201 bp
Download/View Sequence/Evidence
Links
Model
RefSeq RNA
Contig
Genes_seq
Symbol O
HFE
OMIM HGNC sv pr dl ev mm hm sts CCT


Human genome overview page (Build 36.1)
Human genome overview page (Build 35.1)
Map Viewer Home
Map Viewer Help
Human Maps Help
FTP
Data As Table View
Maps & Options
Compress Map
Region Shown:
26,200,858
26,201,201
Go
out
zoom
Ideogram
default
master
Chromosome: 1 2 3 4 5 [6] 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X Y MT
Query: BLAST: (276 letters) [clear]
Color Key for Alignment Scores: <40 40-50 50-80 80-200 >=200
Master Map: Genes On Sequence
Summary of Maps
Maps & Options
Region Displayed: 26,200,858-26,201,201 bp
Download/View Sequence/Evidence
Links
Model
RefSeq RNA
Contig
Genes_seq
Symbol O
MapViewer
Recenter
Zoom in x2
Zoom in x4
Zoom in x8
Zoom out x2
Show 10M
Show 1M
Show 100K
OMIM HGNC sv pr dl ev mm hm sts CCT

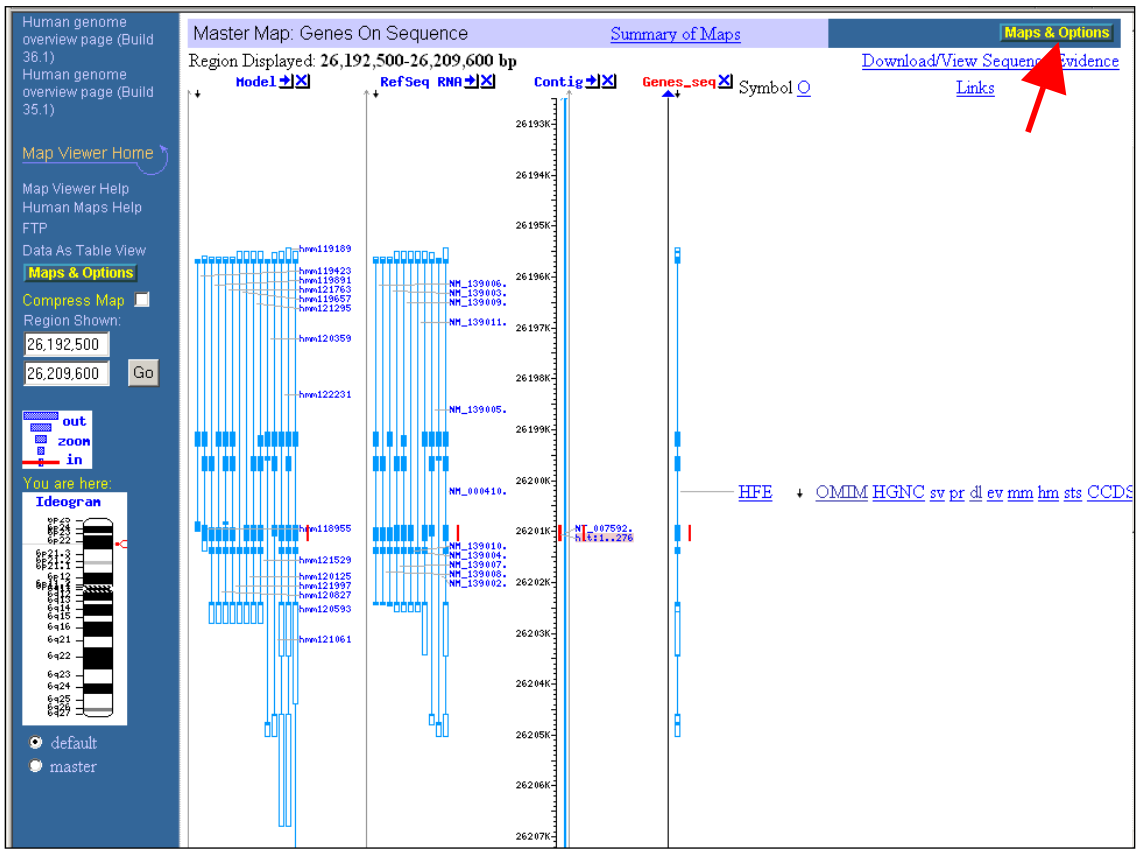

Human genome overview page (Build 36.1)
Human genome overview page (Build 35.1)
Map Viewer Home
Map Viewer Help
Human Maps Help
FTP
Data As Table View
Maps & Options
Compress Map
Region Shown:
26,192,500
26,209,600
Go
out
zoom
in
You are here:
Ideogram
default
master
Master Map: Genes On Sequence
Summary of Maps
Maps & Options
Region Displayed: 26,192,500-26,209,600 bp
Download/View Sequence/Evidence
Links
Model
RefSeq RNA
Contig
Genes_seq
Symbol O
HFL OMIM HGNC sv pr dl ev mm hm sts CCDS

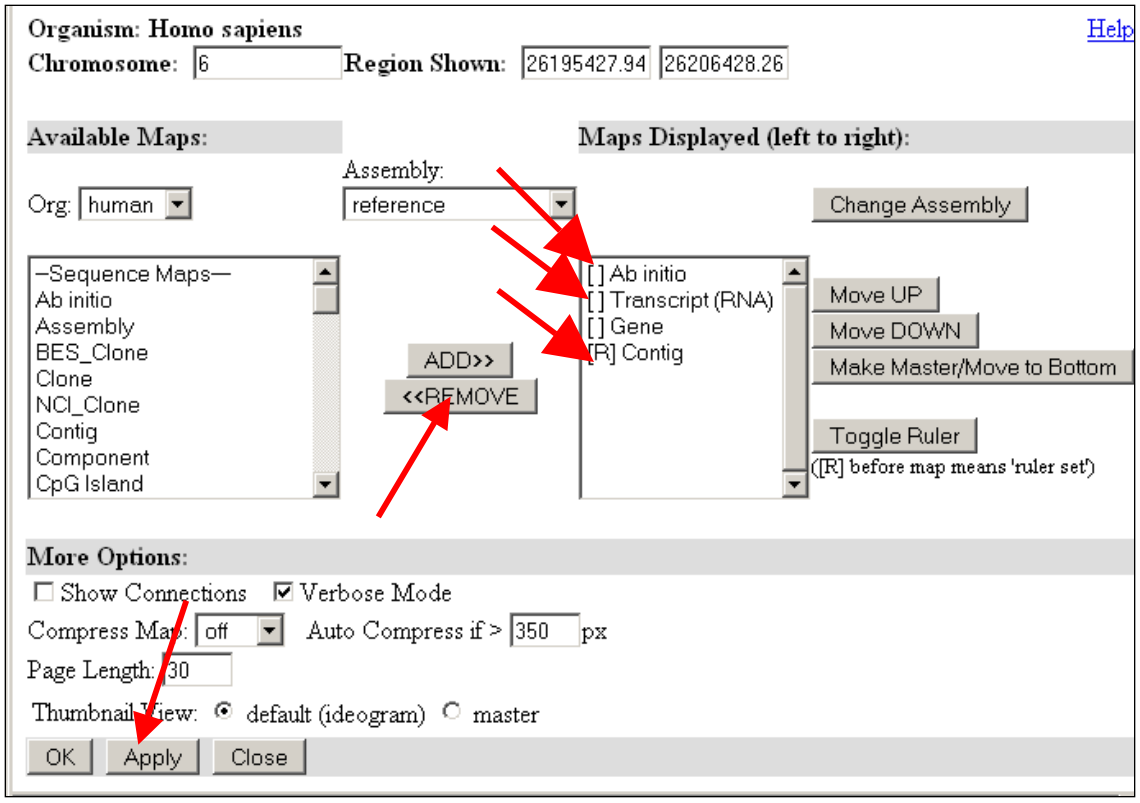

Homo sapiens (Build 36.1)
Region to retrieve (in chromosome coordinates):
Chromosome: 6
Strand: plus
from: 26195427
adjust by: -0K
to: 26205038
adjust by: +0K
Change Region/Strand
Sequence Format: FASTA
This chromosome region corresponds to the contig region(s):
Contig start stop strand
NT_007592.14 16945699 16955310 + Display Save to Disk View Evidence ModelMaker

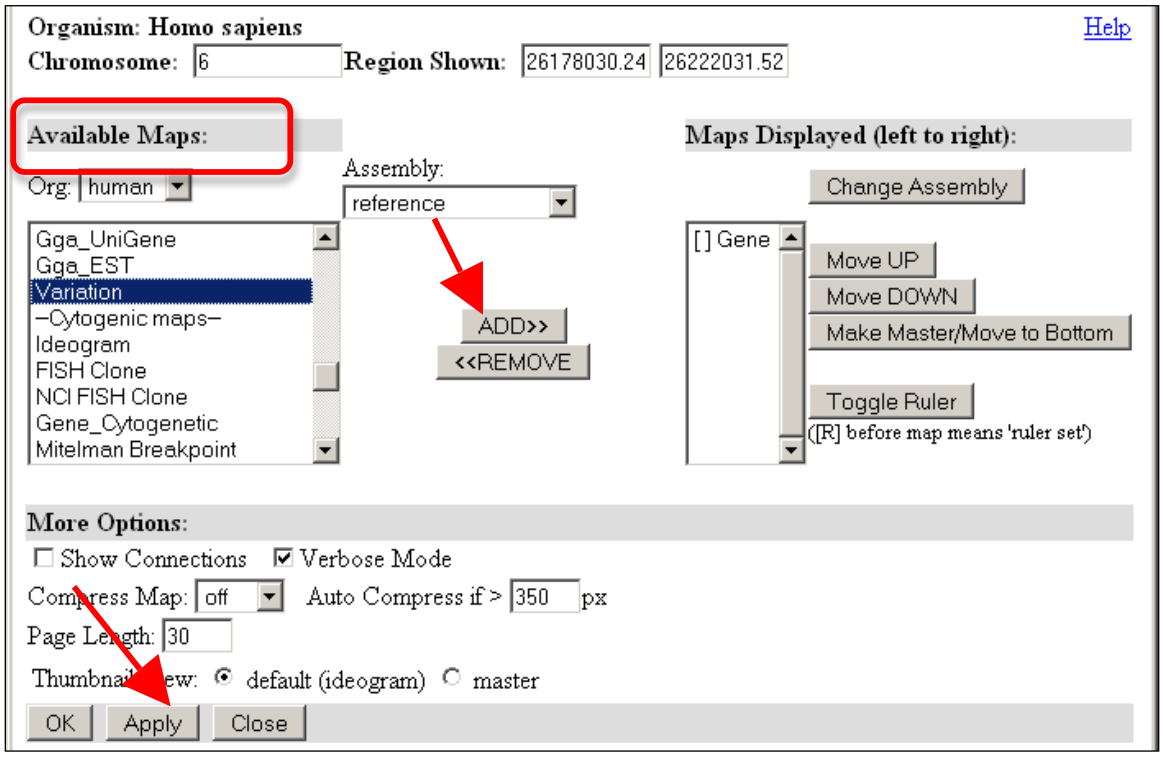


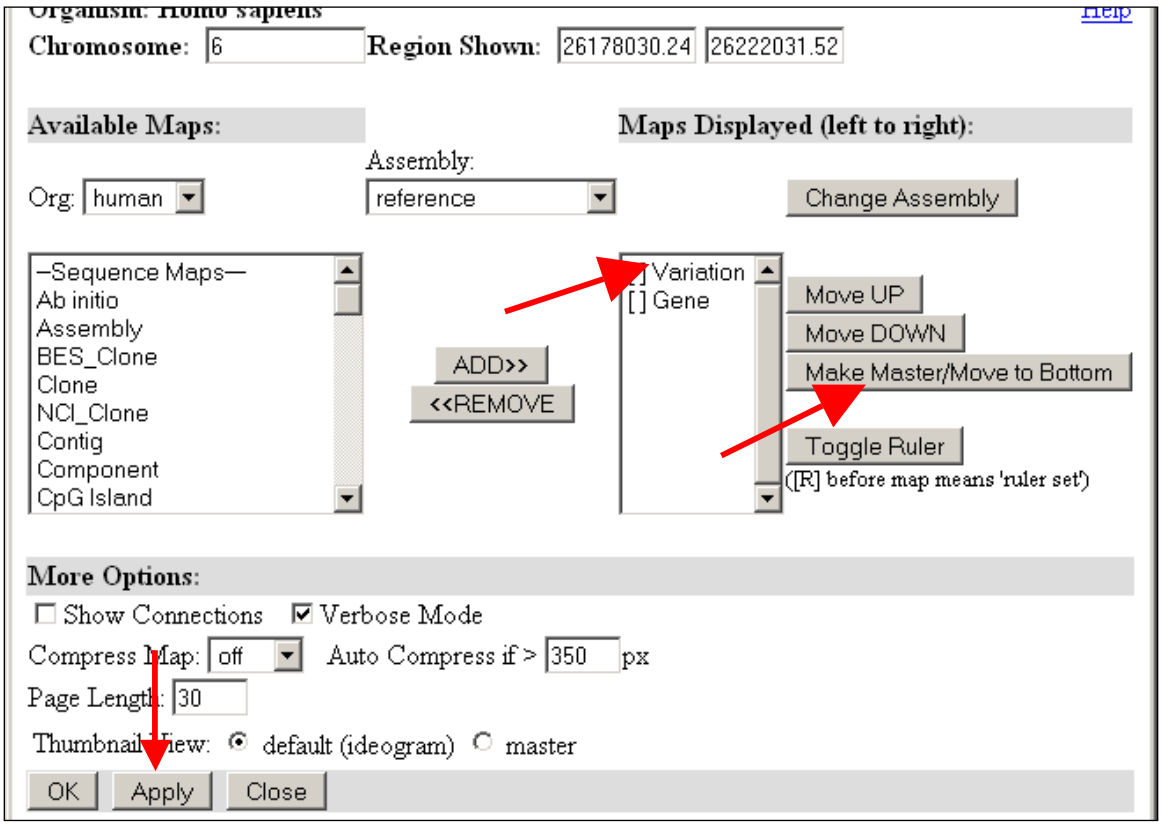
Result: The query EST is expressed by the HFE gene.

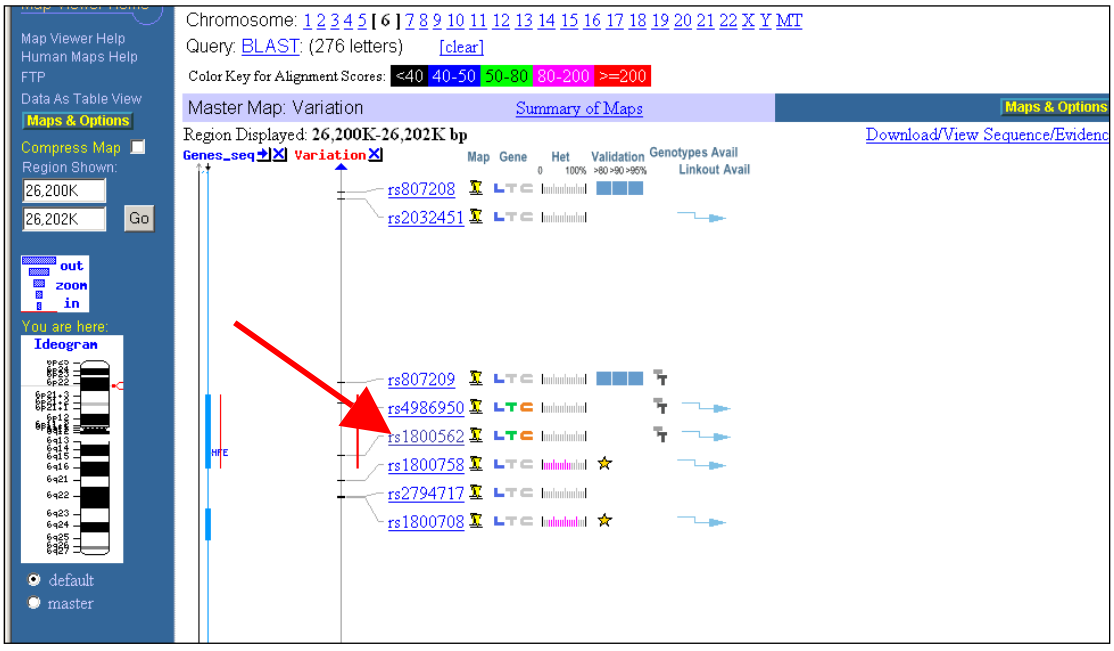
Step 3: Determine whether the EST contains any known SNPs

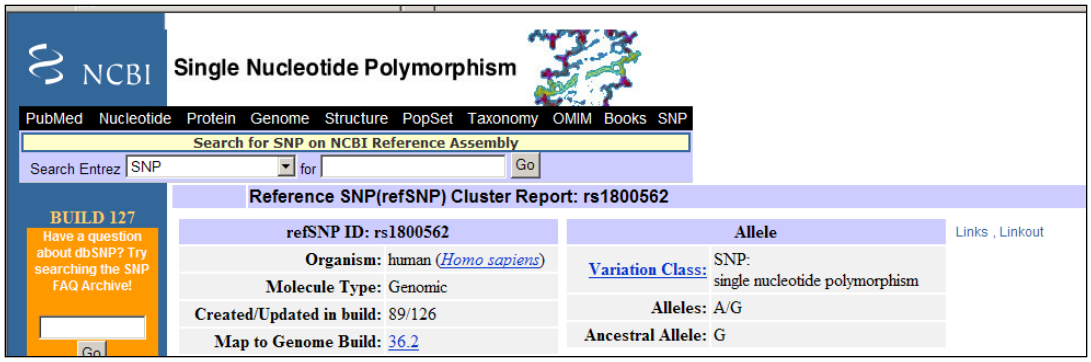


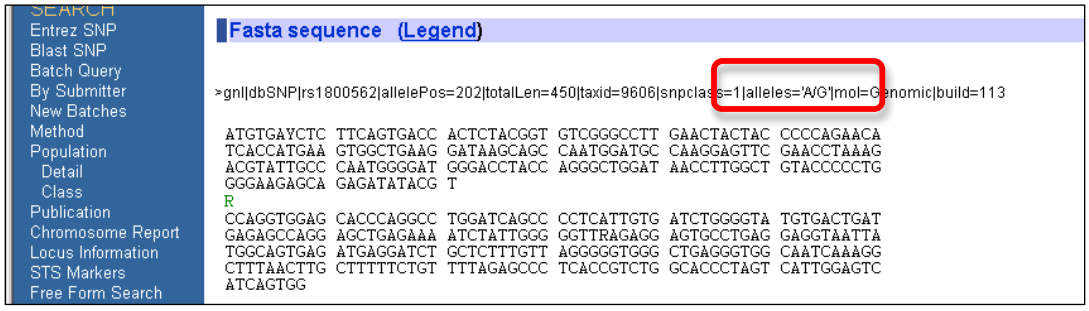

Organism: Homo sapiens
Help
Chromosome: 6
Region Shown: 26195427.94 26206428.26
Available Maps:
Maps Displayed (left to right):
Assembly:
Org: human
reference
Change Assembly
–Sequence Maps–
Ab initio
Assembly
BES_Clone
Clone
NCI_Clone
Contig
Component
CpG Island
ADD>>
<<REMOVE
[ ] Ab initio
[ ] Transcript (RNA)
[ ] Gene
[R] Contig
Move UP
Move DOWN
Make Master/Move to Bottom
Toggle Ruler
([R] before map means 'ruler set')
More Options:
Show Connections
Verbose Mode
Compress Map: off
Auto Compress if > 350 px
Page Length: 30
Thumbnail View: default (ideogram) master
OK
Apply
Close

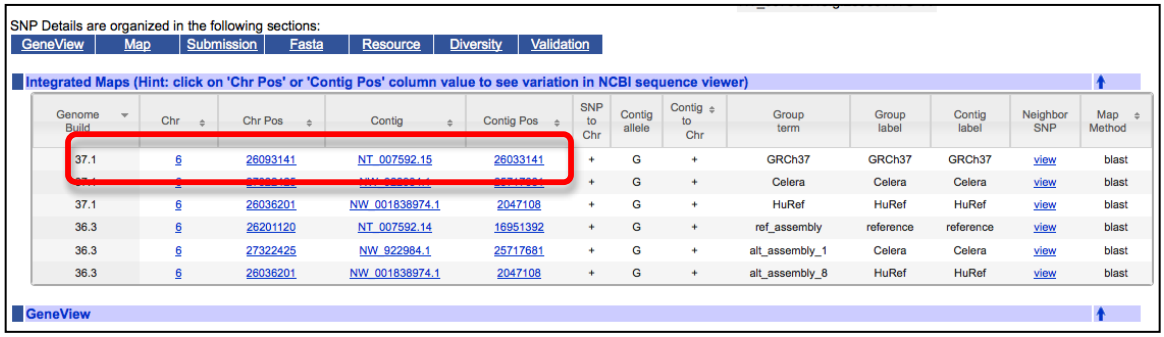

Organism: Homo sapiens
Help
Chromosome: 6
Region Shown: 26178030.24 26222031.52
Available Maps:
Maps Displayed (left to right):
Assembly:
Org: human
reference
Change Assembly
Gga_UniGene
Gga_EST
Variation
–Cytogenic maps–
Ideogram
FISH Clone
NCI FISH Clone
Gene_Cytogenetic
Mitelman Breakpoint
ADD>>
<<REMOVE
[ ] Gene
Move UP
Move DOWN
Make Master/Move to Bottom
Toggle Ruler
([R] before map means 'ruler set')
More Options:
Show Connections
Verbose Mode
Compress Map: off
Auto Compress if > 350 px
Page Length: 30
Thumbnail View: default (ideogram) master
OK
Apply
Close

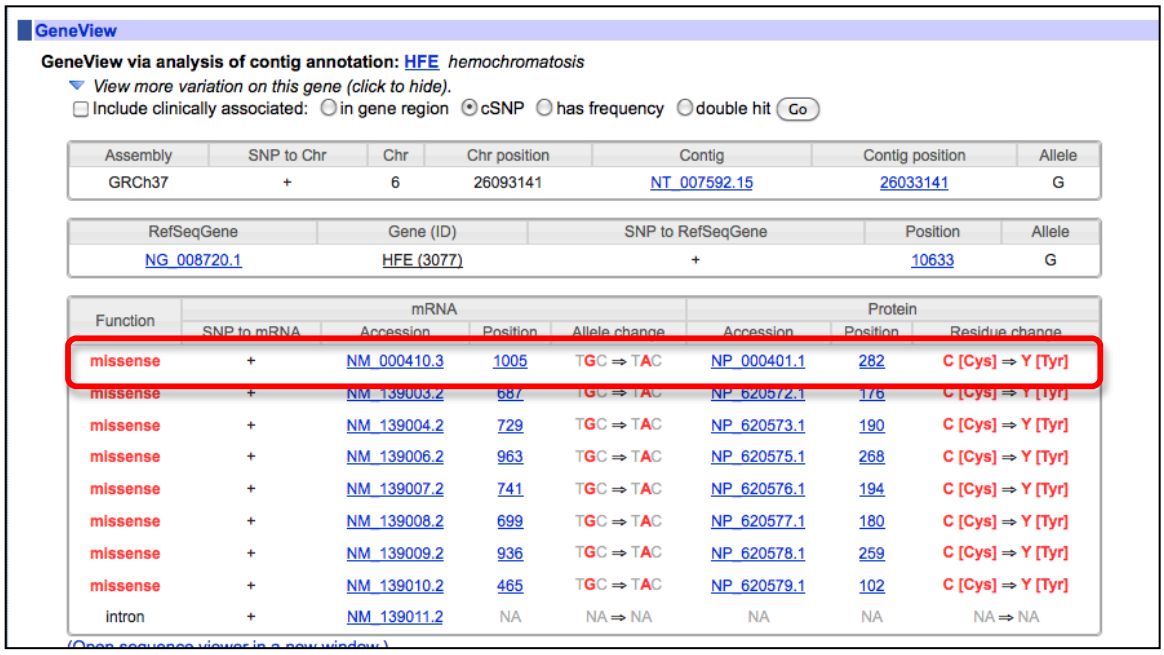

Chromosome: 6
Region Shown: 26178030.24 26222031.52
Available Maps:
Maps Displayed (left to right):
Assembly:
Org: human
reference
Change Assembly
–Sequence Maps–
Ab initio
Assembly
BES_Clone
Clone
NCI_Clone
Contig
Component
CpG Island
ADD>>
<<REMOVE
[ ] Variation
[ ] Gene
Move UP
Move DOWN
Make Master/Move to Bottom
Toggle Ruler
([R] before map means 'ruler set')
More Options:
Show Connections
Verbose Mode
Compress Map: off
Auto Compress if > 350 px
Page Length: 30
Thumbnail View: default (ideogram) master
OK
Apply
Close


Map Viewer Help
Human Maps Help
FTP
Data As Table View
Maps & Options
Compress Map
Region Shown:
26,200K
26,202K
Go
out
zoom
in
You are here:
Ideogram
default
master
Chromosome: 1 2 3 4 5 [6] 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X Y MT
Query: BLAST: (276 letters) [clear]
Color Key for Alignment Scores: <40 40-50 50-80 80-200 >=200
Master Map: Variation
Summary of Maps
Maps & Options
Region Displayed: 26,200K-26,202K bp
Download/View Sequence/Evidenc
Genes_seq
Variation
Map Gene Het Validation Genotypes Avail
Linkout Avail
rs807208
rs2032451
rs807209
rs4986950
rs1800562
rs1800758
rs2794717
rs1800708


NCBI
Single Nucleotide Polymorphism
PubMed Nucleotide Protein Genome Structure PopSet Taxonomy OMIM Books SNP
Search for SNP on NCBI Reference Assembly
Search Entrez SNP for Go
BUILD 127
Have a question about dbSNP? Try searching the SNP FAQ Archive!
Go
Reference SNP(refSNP) Cluster Report: rs1800562
refSNP ID: rs1800562
Organism: human (Homo sapiens)
Molecule Type: Genomic
Created/Updated in build: 89/126
Map to Genome Build: 36.2
Allele
Variation Class: SNP: single nucleotide polymorphism
Alleles: A/G
Ancestral Allele: G
Links , Linkout

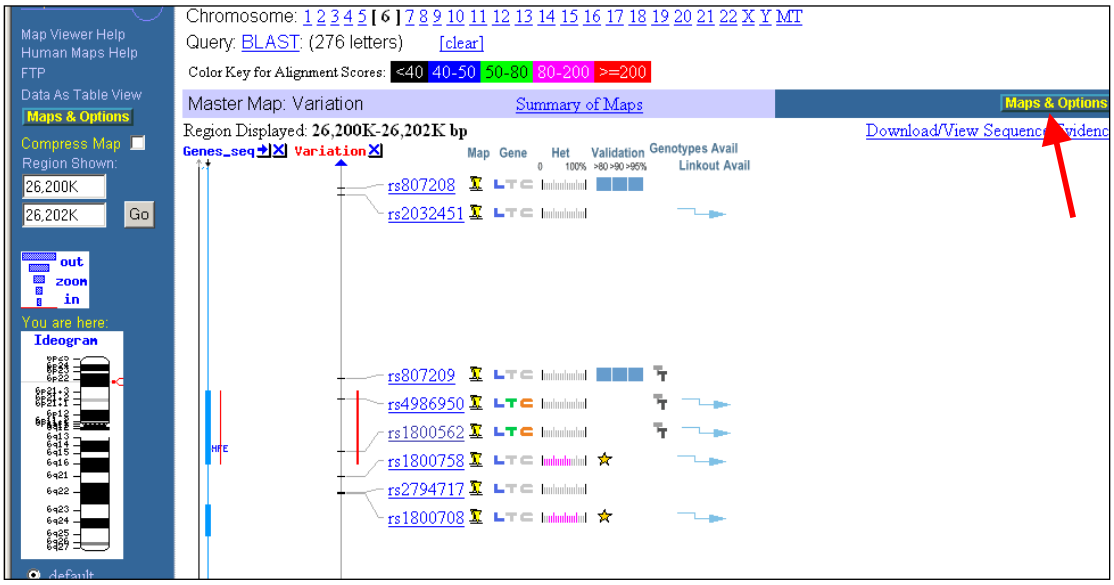

SEARCH
Entrez SNP
Blast SNP
Batch Query
By Submitter
New Batches
Method
Population
Detail
Class
Publication
Chromosome Report
Locus Information
STS Markers
Free Form Search
Fasta sequence (Legend)
>gnl|dbSNP|rs1800562|allelePos=202|totalLen=450|taxid=9606|snpclass=1|alleles='A/G'|mol=Genomic|build=113
ATGTGAYCTC TTCAGTGACC ACTCTACGGT GTCGGGCCTT GAACTACTAC CCCCAGAACA
TCACCATGAA GTGGCTGAAG GATAAGCAGC CAATGGATGC CAAGGAGTTC GAACCTAAAG
ACGTATTGCC CAATGGGGAT GGGACCTACC AGGGCTGGAT AACCTTGGCT GTACCCCCTG
GGGAAGAGCA GAGATATACG T
R
CCAGGTGGAG CACCCAGGCC TGGATCAGCC CCTCATTGTG ATCTGGGGTA TGTGACTGAT
GAGAGCCAGG AGCTGAGAAA ATCTATTGGG GGTTRAGAGG AGTGCCTGAG GAGGTAATTA
TGGCAGTGAG ATGAGGATCT GCTCTTTGTT AGGGGGTGGG CTGAGGGTGG CAATCAAAGG
CTTTAACTTG CTTTTTCTGT TTTAGAGCCC TCACCGTCTG GCACCCTAGT CATTGGAGTC
ATCAGTGG

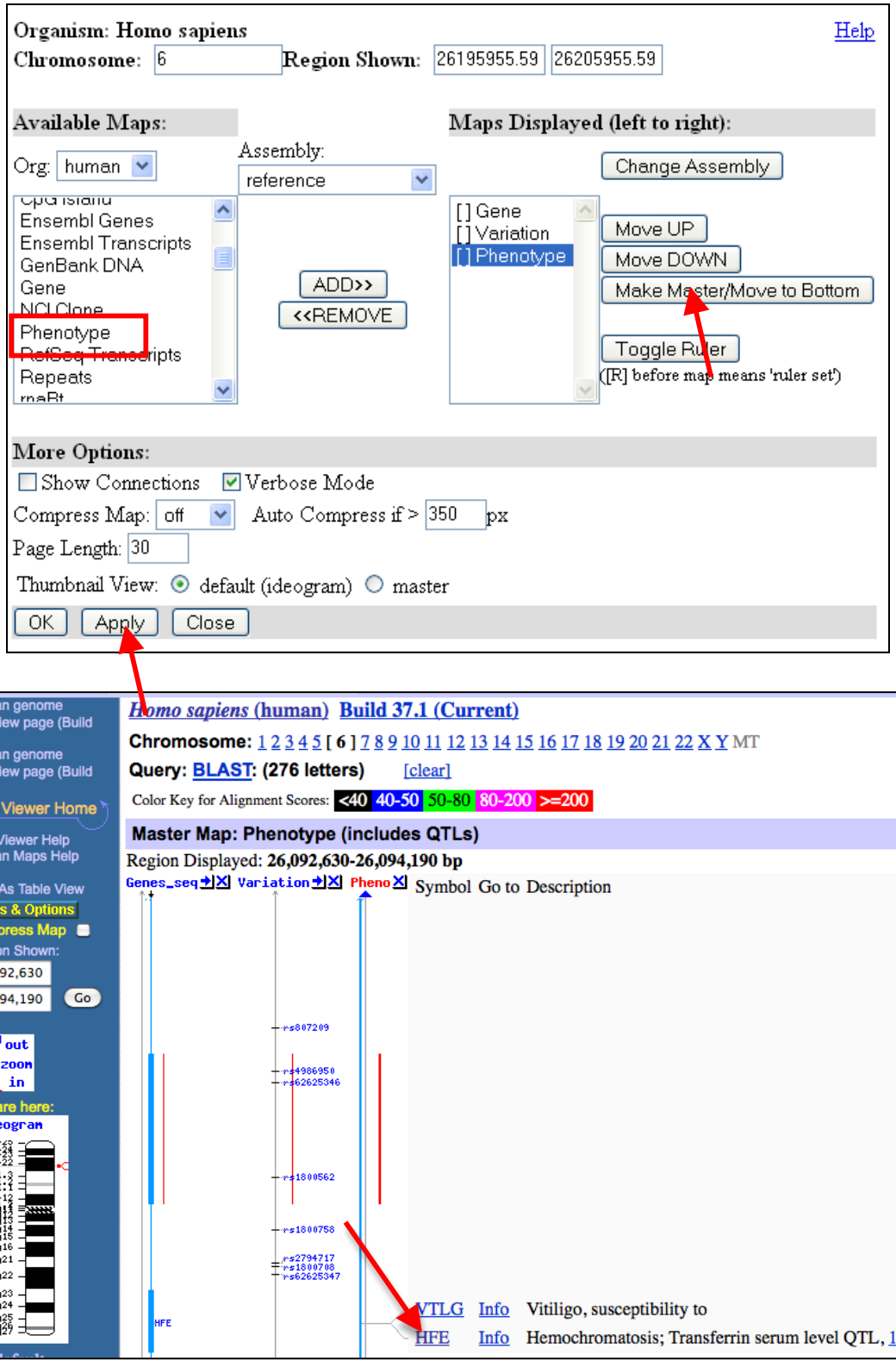

SNP Details are organized in the following sections:
GeneView Map Submission Fasta Resource Diversity Validation
Integrated Maps (Hint: click on 'Chr Pos' or 'Contig Pos' column value to see variation in NCBI sequence viewer)
Genome Build | Chr | Chr Pos | Contig | Contig Pos | SNP to Chr | Contig allele | Contig to Chr | Group term | Group label | Contig label | Neighbor SNP | Map Method
37.1 | 6 | 26093141 | NT_007592.15 | 26033141 | + | G | + | GRCh37 | GRCh37 | GRCh37 | view | blast
37.1 | 6 | 26036201 | NW_001838974.1 | 2047108 | + | G | + | HuRef | HuRef | HuRef | view | blast
36.3 | 6 | 26201120 | NT_007592.14 | 16951392 | + | G | + | ref_assembly | reference | reference | view | blast
36.3 | 6 | 27322425 | NW_922984.1 | 25717681 | + | G | + | alt_assembly_1 | Celera | Celera | view | blast
36.3 | 6 | 26036201 | NW_001838974.1 | 2047108 | + | G | + | alt_assembly_8 | HuRef | HuRef | view | blast
GeneView

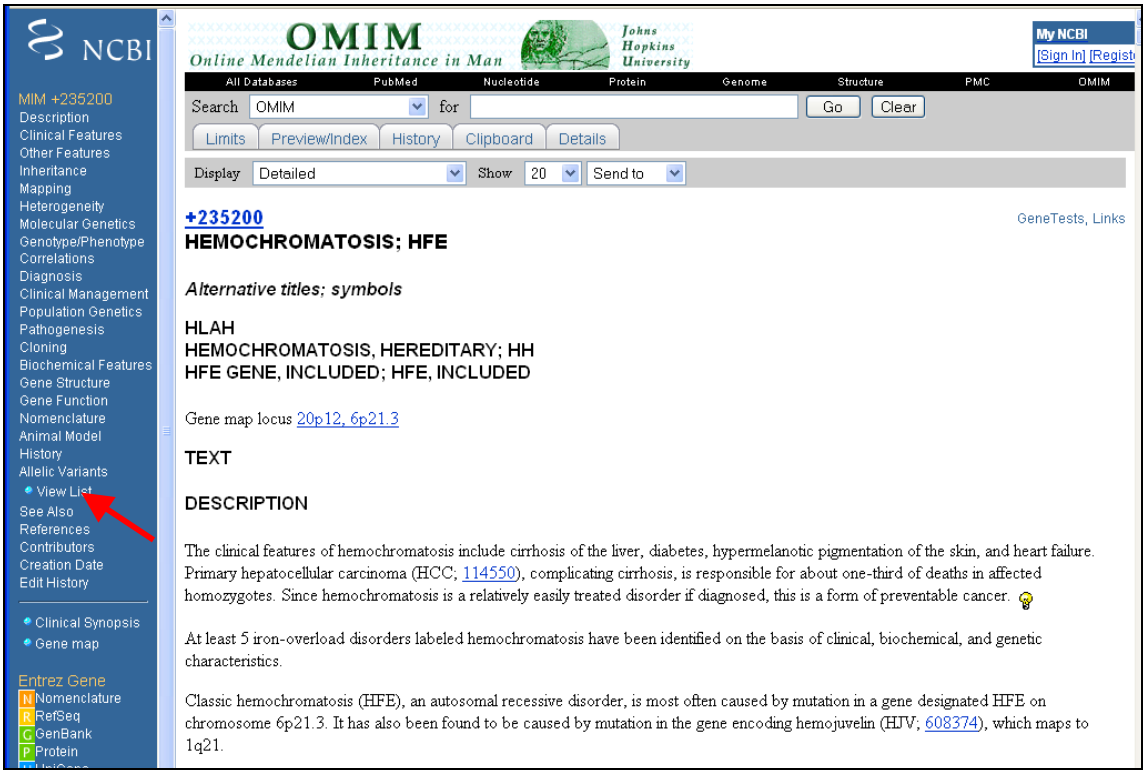


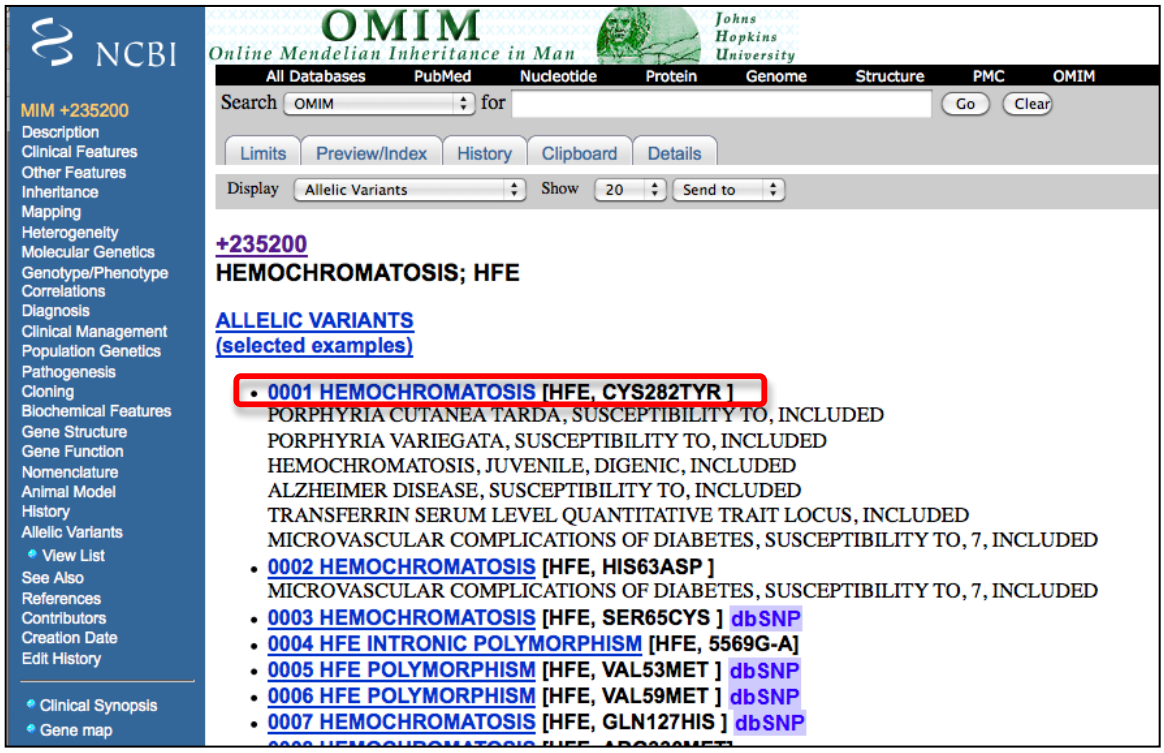
Result: The EST sequence contains a known SNP (G/A with respect to the nucleotide 26033141 on contig NT_007592.15).

Step 4: Determine whether a mutant HFE gene causes a phenotype




Organism: Homo sapiens
Help
Chromosome: 6
Region Shown: 26195955.59 26205955.59
Available Maps:
Maps Displayed (left to right):
Org: human
Assembly:
reference
Change Assembly
Ensembl Genes
Ensembl Transcripts
GenBank DNA
Gene
Phenotype
Repeats
ADD>>
<<REMOVE
[ ] Gene
[ ] Variation
[ ] Phenotype
Move UP
Move DOWN
Make Master/Move to Bottom
Toggle Ruler
([R] before map means 'ruler set')
More Options:
Show Connections
Verbose Mode
Compress Map: off
Auto Compress if > 350 px
Page Length: 30
Thumbnail View: default (ideogram) master
OK
Apply
Close
Homo sapiens (human) Build 37.1 (Current)
Chromosome: 1 2 3 4 5 [6] 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X Y MT
Query: BLAST: (276 letters) [clear]
Color Key for Alignment Scores: <40 40-50 50-80 80-200 >=200
Master Map: Phenotype (includes QTLs)
Region Displayed: 26,092,630-26,094,190 bp
Genes_seq Variation Pheno Symbol Go to Description
rs807209
rs4986950
rs62625346
rs1800562
rs1800758
rs2794717
rs1800708
rs62625347
HFE
VTLG Info Vitiligo, susceptibility to
HFE Info Hemochromatosis; Transferrin serum level QTL, 190000;
Human genome overview page (Build 37.1)
Human genome overview page (Build 36.3)
Map Viewer Home
Map Viewer Help
Human Maps Help
FTP
Data As Table View
Maps & Options
Compress Map
Region Shown:
26,092,630
26,094,190
Go
out
zoom
in
You are here:
Ideogram





Result: Mutations in the HFE gene are associated with hemochromatosis disease

Problem 2:

A laboratory has generated an EST library from a sickle cell anemia patient and wants to identify the gene(s) causing the phenotype. Sickle cell anemia is a disease in which the red blood cells are curved in shape, and which causes pain and fever.

Outline:

We will follow these steps to solve the problem:
1. Compare an EST from a sickle cell anemia patient to the human genome (using BLAST).
2. Identify the gene(s) aligning with the EST and download their sequences (using Map Viewer).
3. Identify whether the EST contains any known nucleotide variations (single nucleotide polymorphisms) (using dbSNP).
4. Determine whether a mutant form of the gene is known to cause a phenotype (using OMIM).

Step 1. Compare ESTs to the human genome (using BLAST):

One way to identify the genes expressing the ESTs is to compare the EST sequence with the human genome assembly and the genes annotated on it. To access the specialized BLAST page for searching against the human genome assembly, click on

BLAST (human genome)

Copy the EST sequence provided below. Paste it in the query box of the BLAST page and select the "genome (reference only)" database. Start the search by clicking on the "Begin Search" button.

ACATTTGCTTCTGACACAACTGTGTTCACTAGCAACCTCAAACAGACACCAT
GGTGCATCTGACTCCTGTGGAGAAGTCTGCCGTTACTGCCCTGTGGGGCAA
GGTGAACGTGGATGAAGTTGGTGGTGAGGCCCTGGGCAGGCTGCTGGTGG
TCTACCCTTGGACCCAGAGGTTCTTTGAGTCCTTTGGGGATCTGTCCACTCC
TGATGCTGTTATGGGCAACCCTAAGGT

Name the chromosome and the contig that we get as a BLAST hit. Note that the similarity is on the minus strand of genome. Is the EST sequence 100% identical to the genomic sequence? Note the nucleotide difference and position between the two sequences.

Next, open another window and use Align 2 or more sequences (linked on the BLAST homepage) and align the EST sequence to NC_000011, the chromosome record for chromosome 11. The results will appear similar, but note the position of the mismatch relative to the chromosome. The coordinates are slightly different between the contig and chromosome records. You will need the chromosome coordinates for the next part of the exercise. Now return to the previous window where you ran the human genome BLAST query.

Step 2. Identify the gene(s) expressing the EST and download their sequences:

To visualize the BLAST hit on the genome using Map Viewer, click on the "Genome View" button at the top of the results page, then on the Map element "NT_009237". Currently, 4 maps should be displayed (Contig, Model, RNA and Gene_seq). Zoom out 2 or 4 times by clicking on right most contig map and selecting the appropriate option.
The best BLAST hits, indicated by the red bars, are in the region of two exons of the HBB gene annotated on the human genome. Make the Gene_seq map a master map by clicking on the arrow at the top of the map. Note that the gene is annotated on the minus strand. To display the entire HBB gene sequence, click on the "dl" link, choose minus strand from the pull down menu, click on "Change Region/Strand" and display the sequence by clicking on "Display". Copy the sequence and paste it in the area provided below. We will use it later to obtain the exon-intron structure. You can adjust the nucleotide locations to download the upstream or downstream sequence by using the "adjust by" and "Change Region/Strand" option.

Step 3. Determine whether the EST contains known SNPs:

Go back to the Map Viewer report. Click on the Maps and Options link. Remove all the maps except the Gene_seq map by selecting the map under the Maps Displayed menu and clicking on Remove. Now add the variation map from the Available maps menu (by selecting the map and clicking on Add). Make the Variation map as the master map by selecting it and clicking the Make Master/Move to Bottom option. Then click on Apply. Now two maps are displayed, Variation (it's the rightmost and the master map) and Gene_seq. The master map provides detailed information for the map features, in this case SNPs. ". (The Map Viewer exercise describes the usage of the Map Viewer in detail.) Zoom in on the blast hit area (red bar). There are numerous SNPs in the area, one of them is rs334. To see it, type in the exact genomic coordinates in the region shown box (5,248,232) Click on the link for the SNP. There is an A/T SNP is at the nucleotide position 5188232 on the contig NT_009237 as mentioned under Fasta sequence and Integrated maps. Is this the same nucleotide variation found in the BLAST result in Step 1?

Step 4. Determine whether the mutant HBB gene causes a phenotype:

Go back to the Map Viewer report. Go back to the Map Viewer report and add the Phenotype map as the master map using the Maps and Options menu. Click on the HBB link that leads to the hemoglobin beta record (141900) in the OMIM database. The record details how mutations in the HBB gene are associated with a phenotype, sickle cell anemia. As mentioned in the report, the allelic variants are listed for the mature HBB protein which lacks initiator methionine. Click on the Allelic Variant "View list" to get information about mutant proteins from patients. Is Glu6Val variant mentioned in the list? Which phenotype does it cause?

Summary:

This exercise describes steps to identify the gene expressing the ESTs obtained from a sickle cell anemia patient, download the gene sequence, identify known SNPs in the gene and find SNP-associated phenotypes.

Step 1: The query EST sequence was found to align to contig NT_009237.17 on chromosome 11 with one nucleotide difference (T to A with respect to the nucleotide 5188232 on the contig).
Step 2: The query EST was found to be expressed by the HBB gene.
Step 3: The query EST sequence contains a known SNP (T/A with respect to the nucleotide 5188232 on contig NT_009237.17).
Step 4: Mutations in the HBB gene are associated with sickle cell anemia.